The Probabilistic Inference of Unknown Data in Phylogenetic Analysis

André NEL

Laboratoire d'Entomologie, Muséum national d'Histoire naturelle,
45, rue Buffon, 75005 Paris, France

ABSTRACT

Unknown characters and attributes are inferred in phylogenies using a probabilistic method. The probability of the position of fossil taxa having uncertain relationships, because of the lack of unambiguous synapomorphies, can be calculated using a similar method. These methods allow a better definition of the limits of actualism in Paleontology and can be applied to palaeoclimatic and palaeoenvironment studies.

RÉSUMÉ

Inférence probabiliste de données inconnues en analyse phylogénétique

Il est proposé une méthode probabiliste d'inférence phylogénétique de caractères et attributs d'état inconnu ainsi qu'une méthode analogue de calcul de la probabilité de la position de taxa fossiles, *a priori* incertaine par manque de synapomorphies non ambiguës. Ces méthodes permettent de quantifier les hypothèses d'inférence basées sur les liens de parenté des taxa fossiles et ainsi de définir les limites de l'utilisation du principe de l'actualisme en paléontologie, en particulier pour la paléoclimatologie et les analyses paléoenvironnementales.

INTRODUCTION

One of the main problems in Paleontology is the reconstruction of the palaeobiotas and palaeoclimates by comparison between fossil and recent taxa. One can use the actualist method which extends the Recent biological and ecological data to the past. FURON (1964) gave an example which represents a good summary of the use of actualism: "Les grands Foraminifères [...] vivaient dans des mers chaudes. [...] mais c'est évidemment la répartition des récifs coralliens qui nous donne les renseignements les plus précis". But incorrect use of actualism could be misleading and several examples reveal particularly unreliable for the study of very old palaeoenvironments. After BRYANT & RUSSELL (1992), there are two different methods of inference of Recent data to the past: a) after the morpho-functional inference theory, a fossil organ identical to a Recent one had a similar function ; b) after the phylogenetic inference theory,

a fossil taxon related to a Recent one had a similar biota, under a similar climate and environment. Being limited to functional data, the first type of inference is often useless for palaeoclimatic analysis. Furthermore, confusions between the two types of inference are frequent and are leading to abusive conclusions, supported by weak evidences. Because of the lack of clear synapomorphies with Recent groups, the phylogenetic positions of fossil taxa are frequently very uncertain. Evenly, one would be tempted to use phenetic methods for classifying these taxa. Either BRYANT & RUSSELL's method (*loc. cit.*) only give qualitative phylogenetic inferences. As methods of quantification of data inferences and of phylogenetic positions of taxa are lacking, we have attempted to define them with a probabilistic theory based on cladistic analysis. We did not attempted to use a maximum likelihood method of analysis because the informations concerning the various probabilities for the ancestor state, change of states along the branches, etc. (DARLU & TASSY, 1993), are never available for the inference of complex palaeoclimates or palaeoenvironmental informations.

DATA INFERENCES (FOSSIL / RECENT)

Using phylogenetic systematics, it is possible to extend informations from Recent data to the fossil record on the basis of the systematic position of the fossil taxon. Informations (of palaeoclimatic and palaeoenvironmental types between others) are then to be considered as attributes (*sensu* MICKEVICH & WELLER, 1990; DELEPORTE, 1993; GRANDCOLAS, 1993; GRANDCOLAS *et al.*, 1994). An attribute is a trait of extrinsic type. Its primary homology (*sensu* DE PINNA, 1991, before any phylogenetic analysis) is not assessed, but its similarity can be postulated, in order to give the same name to traits of the different taxa. A character is, in phylogenetic analysis, a trait unambiguously homologous in several taxa before any the phylogenetic analysis. The polarization of an attribute is to be made by optimization on the phylogenetic tree. BRYANT & RUSSELL (1992) have established a general method of inference of characters (or attributes), the states of which are unknown for taxa already included in a phylogenetic analysis : if the taxon (fossil or not) is included in a group having an homogeneous state for the considered attribute (or character), the probability for the (fossil) taxon to have an information homologous to the data given by the Recent taxa increases. On the contrary, if the (fossil) taxon is only the sister-group of a Recent group, the inference of information becomes more uncertain and cannot be justified with the sole hypothesis of parsimony. Only an hypothesis of phylogenetic proximity provides support for the hypothesis of inference. This method of inference only tests for presence/absence of a character in one taxon while it is present in other taxa. It does not concern the possible autapomorphies of the fossil taxon for the studied attribute. It is impossible to infer phylogenetically an unknown autapomorphic character for a fossil taxon, on the basis of characters of different type of the nearest Recent relatives. "Phylogenetic inference is conservative" (BRYANT & RUSSELL, 1992).

The method of BRYANT & RUSSELL is based on the outgroup "ascendant" algorithm of MADDISON *et al.* (1984) : the situation of a character (or an attribute) at each internal node of the tree is parsimoniously inferred by the situations at the two immediately adjacent nodes. For a character X with two states "a" and "b", an internal node is labeled "a" if the two immediate adjacent nodes are labeled "a" and "a" or "a" and "a or b". Symmetrical situation occurs with "b". Nodes are labeled "a or b" if the adjacent nodes are labeled "a" and "b", or "a or b" and "a or b". The external node with the missing information is then supposed to share the same state as

the immediately internal adjacent node. This method is problematic for the inference of characters that have been used in the construction of the minimal tree, using adequate softwares because the computer programs Hennig86 and Paup 3.1.1 trend to affect definite values to the missing entries, even if they do this in very different ways : in Paup, "only those characters that have non-missing values" are supposed to "affect the location of any taxon on the tree" (SWOFFORD, 1990) because Paup assigns to the taxon affected by the missing character the character state that would be most parsimonious given its placement in the tree. Equally parsimonious trees are constructed for the concerned character and then discriminated on the other non missing characters. PLATNICK *et al.* (1991) have tested Hennig86 and Paup and would confirm the assumption of SWOFFORD. They add that Hennig86 trends to attribute global peculiar states to missing data, depending of the tree topology. Paup gives a less resolved solution than Hennig86, because it interpret the missing data in the construction of the minimal trees. Nevertheless, if the missing data are reinterpreted and used in the construction of the minimal tree, it is delicate to test their value on the basis of the same tree. This difficulty does not exist if the concerned character is considered as an attribute, not included in the preliminary construction of the tree and independently tested after this construction.

BRYANT & RUSSELL could not quantify their hypothesis of congruence, but it is possible to calculate the probability of the following event: [the missing information is homologous to the information given by the nearest relative taxa]. A preliminary hypothesis is necessary: the inference of the unknown situations for the F-taxa will not add homoplasies or steps to the general shape of the concerned attribute in the minimal tree. All the situations that do not imply supplementary homoplasies or supplementary steps are then supposed to be equally probable. Following this condition, the probability of the event [the studied taxon has a peculiar state for the studied character] can be calculated by making the ratio of the number of favorable situations by the total number of possible situations.

Bipolar attribute X

Theoretical procedure. X is supposed to have two states "a" and "b". As the polarity of the attribute and its homoplasy rate are completely unknown, we consider that it can equally be in the states "a" or "b" in the root of the tree. Then, the minimal scenarios (with the lowest number of steps) that explain the known distribution of the attribute (excluding the F-taxon) are reconstructed for the two situations "state a for the root" (or "root : a") and "root : b".

Then, two options are possible:

option (1): either these two minimal scenarios concerning the possible situations of the root can either be considered as equally probable, even if one can imply more steps than the other.

option (2): or the minimal scenarios that explain the known distribution of the attribute with a "root : b" is affected of a weight (x); then the minimal scenario for the "root : a" has a weight $(1 - x)$. They are not considered as equally probable; $0 \leq (x) \leq 1$. Option (1) is a peculiar case of option (2), with $(x) = 0.5$. Nevertheless, option (1) corresponds to the minimal *a priori* scenario (equal weight for the two situations of the root).

After, the F-taxon is re-included in the tree, then are only accepted the situations where the alleged state for F can be reconstructed without adding supplementary homoplasies or steps in the tree ("favorable cases"), in order to keep the same minimal lengths for the new trees. This method of inference do not add other hypothesis than the two possible situations in the root of

the tree. Finally, the "favorable cases" are counted, and the probability of the event (F is in the state "a") labeled "p (F : a)" is simply the following ratio:

p (F : a) = (number of cases favorable to the state "a")/(number of favorable cases)

Application. 1) If a taxon F is simply the sister-group of a known taxon A, without further information, the probability for the event [F and A share the same state for a character X] = 0,5, which can be written p (F : a) = 0,5 = p (F : b). A simple information of sister-group relationship does not allow any prospective for unknown characters.

2) If a taxon F is the sister-group of a known taxon A1, [F + A1] being the sister-group of a known taxon A2, and if A1 and A2 share the state "a" of the character, unknown for F (Fig. 1):

BRYANT & RUSSEL's method infers the situation "a" for F.

Using the probabilistic method,
with option (1) for the roots, I have:

$$p (F : a) = 1 \text{ and } p (F : b) = 0 \quad \text{(Fig. 1)}$$

with option (2) for the roots, I have:

$$p (F : a) = [x + (1 - x)] / [x + (1 - x)] = 1$$

Thus the results of the two methods are congruent, in either situations.

3) If the taxon A2 has the contrary state "b":

After BRYANT & RUSSEL'S method, the inferred situation is "a or b" for F.

Using the new probabilistic method,
with option (1) for the roots, we have:

$$p (F : a) = 2/3 \text{ and } p (F : b) = 1/3 \text{ (Fig. 2)}.$$

with option (2) for the roots, we have, as p (F : a) depends on x:

$$p (F : a) (x) = [x + (1 - x)] / [(x + x + (1 - x)] = 1 / (1 + x)$$

As $0 \leq x \leq 1$, p (F : a) (0) = 1 ; p (F : a) (1) = 0.5 ; and p (F : a) (0.5) = 2/3 [as x = 0.5 corresponds to option (1)].

More generally, $0.5 \leq p (F : a) (x) \leq 1$

With the minimal scenario for the roots, p (F : a) = 2/3

The results of the two methods are congruent but we can quantify the alternative "a or b".

4) If we add to the schema of Figure 1 the information (Fig. 3) of an out-group (including one or several taxa) A3 which has a state "b" :

BRYANT & RUSSEL's method infers the situation "a" for F.

With the probabilistic method,
with option (1) for the roots, I have:

$$p (F : a) = 1 \text{ and } p (F : b) = 0$$

with option (2) for the roots, I have:

$$p (F : a) (x) = [x + (1 - x)] / [x + (1 - x)] = 1$$

The two results are clearly congruent. Adding supplementary taxa which would be branched lower in the phylogenetic tree will not change the probabilities for the terminal branch which includes F.
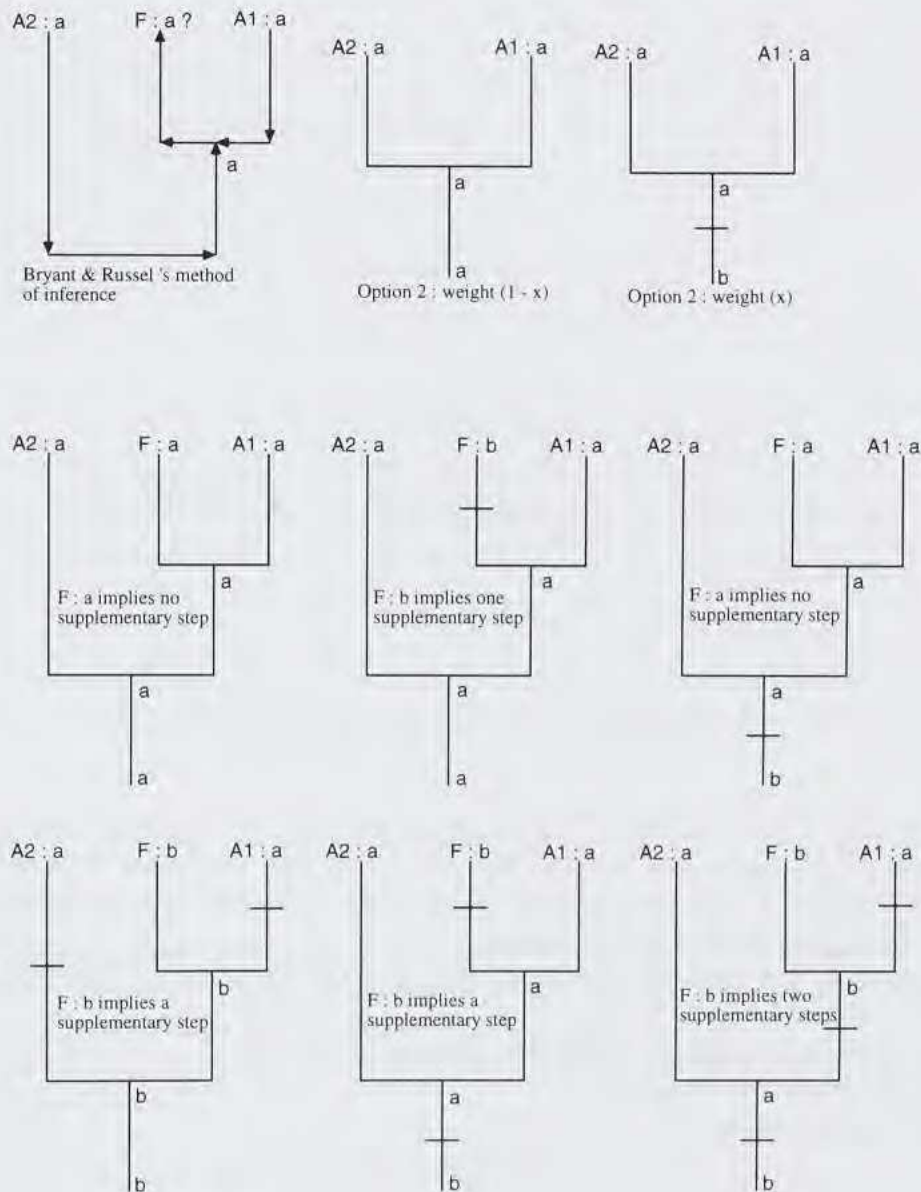
FIG. 1. — p(F : a) = 1 ; other situations also imply supplementary steps.

5) But if the ingroup includes different taxa with different states, or polymorphic taxon(a), the probability will decrease. In Figures 4 and 5, we give two different examples.

First example: If a third taxon A3 with the contrary state "b" is added to the situation of Figure 2 (Fig. 4) as sister group of [(A1 + F) + A2], the probability is p (F : a) = 1/2 for the two
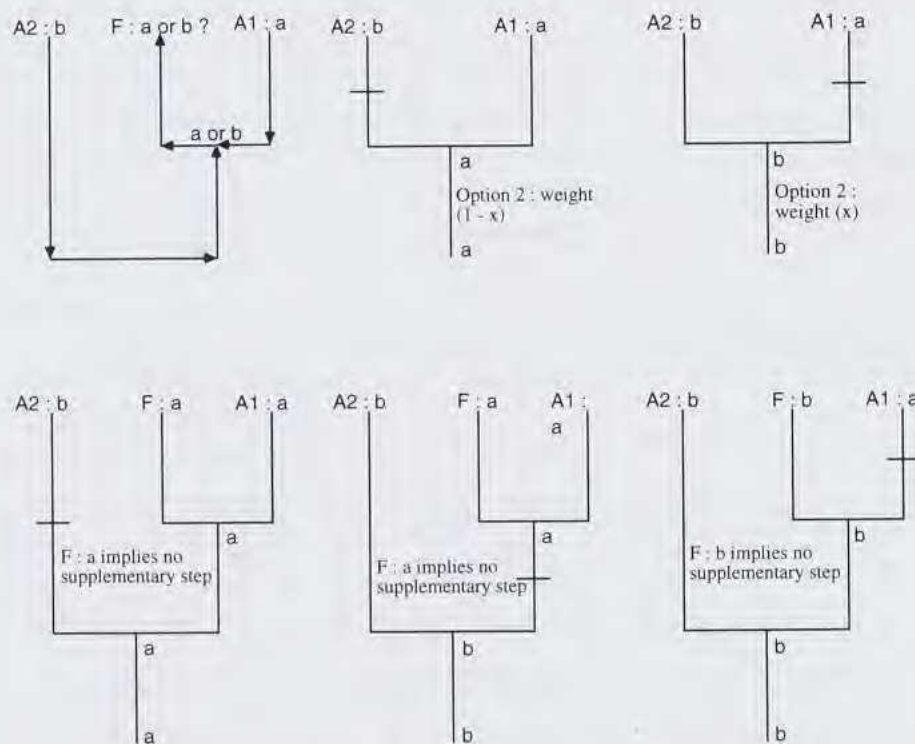
FIG. 2. — p(F : a) = 2/3 ; other situations imply supplementary steps.

options (1) or (2), it has decreased, compared to the result of Figure 2. This result remains congruent with the predictions of BRYANT & RUSSEL's method (inference of "a or b"), but there is a kind of "attraction" of the low branches. Nevertheless, if further taxa having the state "b" are added more basally, the result will not change.

First example: If a third taxon A3 with the contrary state "b" is added (Fig. 5) as sister-group of [(A1 + A2) + F]:

BRYANT & RUSSEL's method gives the inference "b" for F.

With probabilistic method, I have:
with option (1) for the roots:

$$p \, (F : a) = 1/3$$

with option (2) for the roots:

$$p \, (F : a) \, (x) = (1 - x) \, / \, [x + (1 - x) + (1 - x)] = (1 - x) \, / \, (2 - x)$$

p (F : a) (0) = 0.5 ; p (F : a) (1) = 0 ; and p (F : a) (0.5) = 1/3 (minimal scenario for the roots). More generally : $0 \leq p \, (F : a) \, (x) \leq 0.5$, because $0 \leq x \leq 1$ and p (F : a) is a decreasing function of x.

Thus, in this case, the two methods are not congruent, but the probabilistic method agrees with the intuitive assumption of more uncertainty in the inference if the sister taxon (A1 + A2) of F is polymorphic.
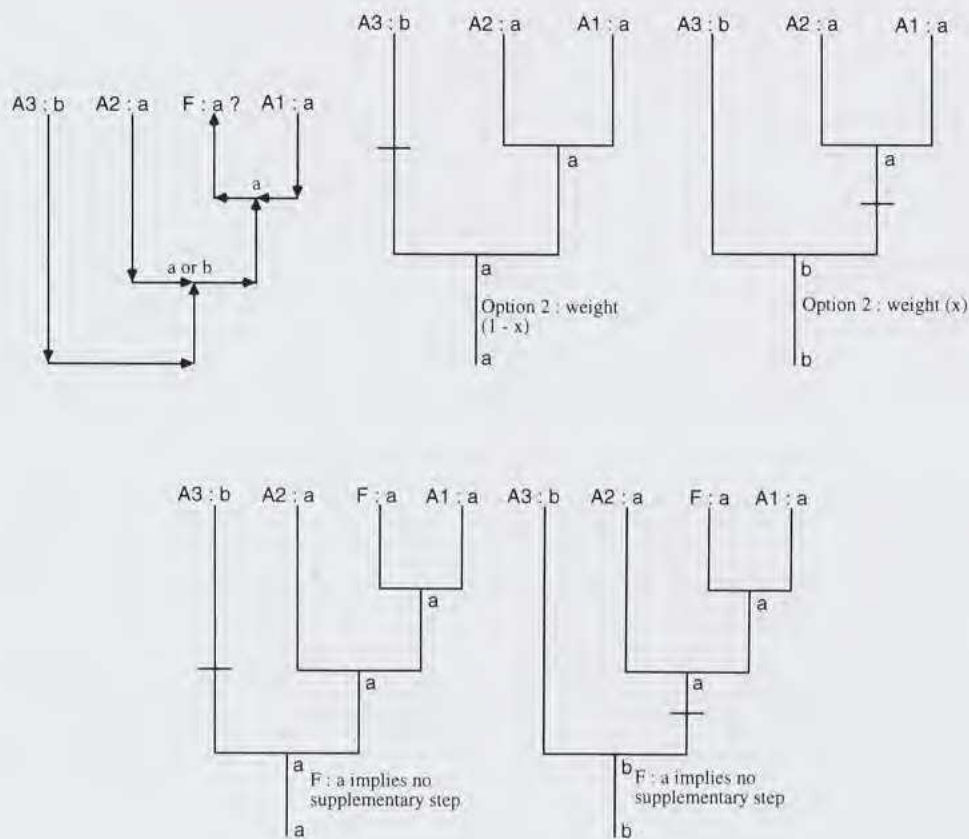
FIG. 3. — p(F : a) = 1.

6) In the case of several taxa with an unknown state for character X (Figure 6), BRYANT & RUSSELL (*loc. cit.*: 410-411, Fig. 4) concluded that no peculiar F-taxon is privileged in the inference of the informations. My study leads to similar results.

In the situation of Figure 6, BRYANT & RUSSEL'S method implies the same inference ("state a") for the two taxa F1 and F2.

The probabilistic method gives the same probabilities for F1 and F2, for the two options (1) or (2), clearly not depending on the position of the fossil taxa in the tree:
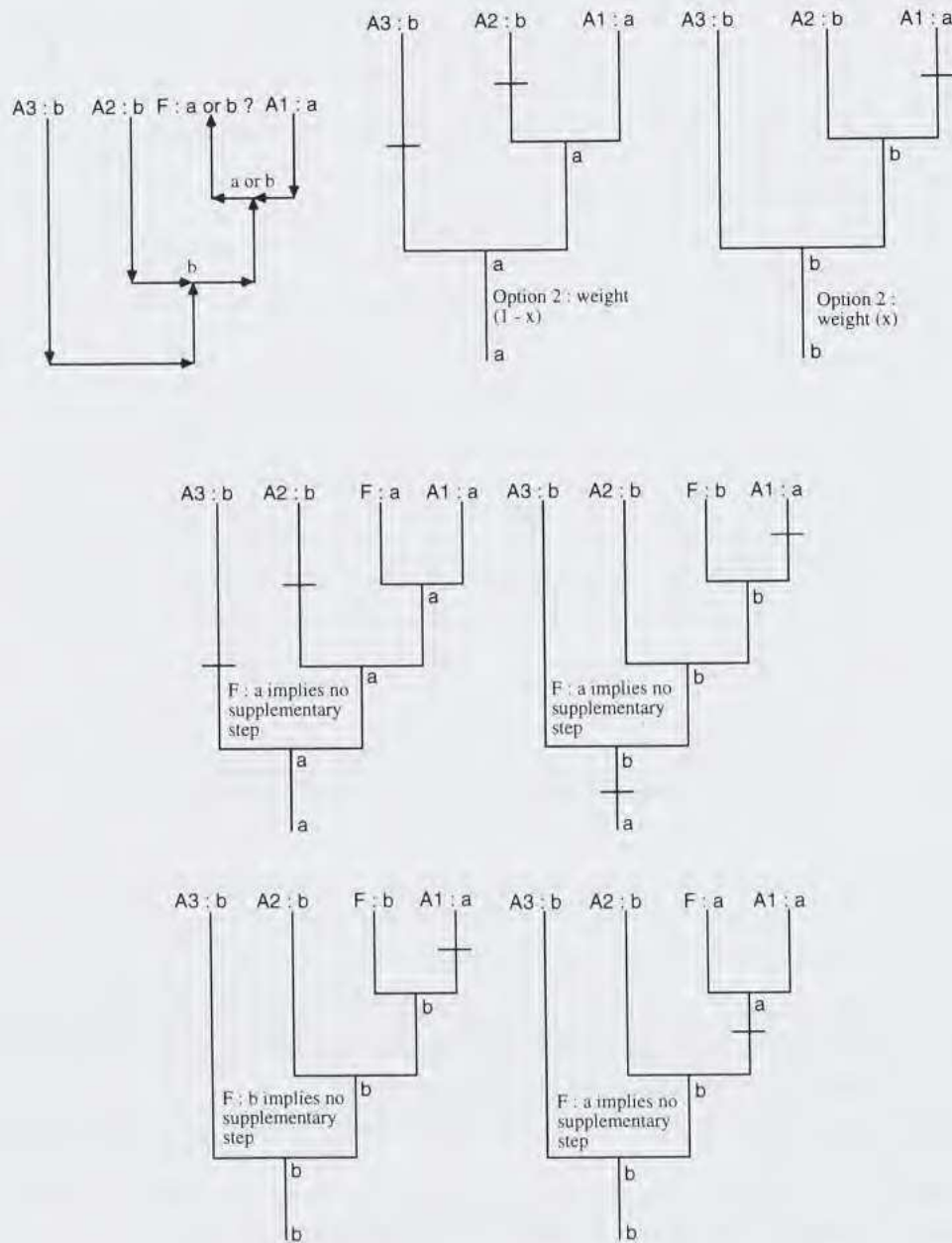
$$p (F1 : a) = 1 = p (F2 : a)$$

Similar results can be obtained with more F-taxa in similar positions.

7) The same calculations have been made in the case of two F-taxa (Fig. 7), but with a situation similar to that of Figure 2:

Bryant & Russel's method gives the same inference of "a or b" for F1 and F2 but it does not add information to the situation of Figure 2.

With probabilistic inference, I have with option (1) for the roots:
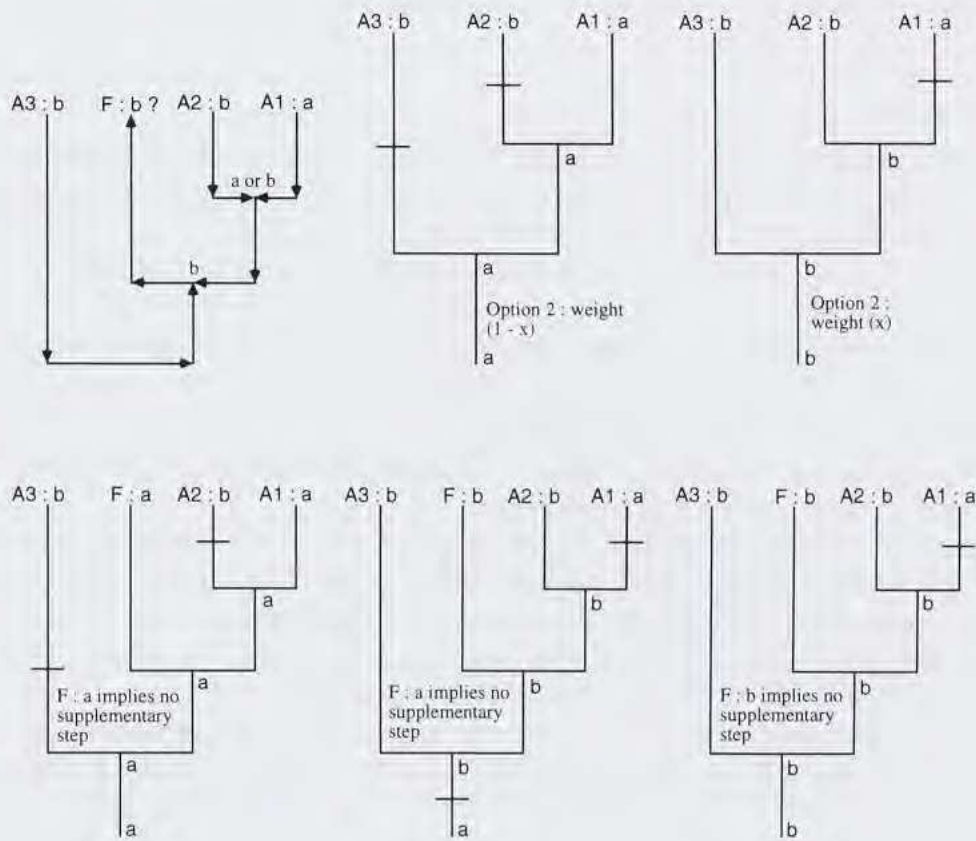
$$p (F1 : a) = 3/4 \text{ but } p (F2 : a) = 1/2$$

FIG. 4. — p(F : a) = 1/2.

with option (2) for the roots:

$$p (F1 : a) (x) = (x + 1) / (2x + 1) \text{ and } p (F2 : a) (x) = 1 / (2x + 1)$$
$$p (F1 : a) (0) = 1 \text{ ; } p (F1 : a) (1) = 2/3 \text{ ; and } p (F1 : a) (0.5) = 3/4$$
$$p (F2 : a) (0) = 1 \text{ ; } p (F2 : a) (1) = 1/3 \text{ ; and } p (F2 : a) (0.5) = 1/2$$

FIG. 5. — p(F : a) = 1/3.

More generally:

$$2/3 \leq p\,(F1 : a)\,(x) \leq 1 \text{ and } 1/3 \leq p\,(F2 : a)\,(x) \leq 1$$

because $0 \leq x \leq 1$; p (F1 : a) and p (F2 : a) are decreasing functions of x.

Thus, in this case, the probabilities are not the same, because of the "attraction" of the low branch. Furthermore, all the probabilities decrease with the number of F-taxa between A2 and A1. This result is congruent with the assumption that the possibilities of the displacements of the positions of the transformations within the tree increase with the number of F-taxa between A1 and A2.

8) With n F-taxa in the same situation as in Figure 7, probabilistic method gives [option (1)]:

$$p\,(F1 : a) = (n+1) / (n+2), \; p\,(F2 : a) = n / (n+2), \; ..., \; p\,(Fi : a) = (n+2-i) / (n+2), \; ..., \text{ and}$$
$$p\,(Fn : a) = 2 / (n+2)$$

If the rank "i" of a F-taxon is supposed constant, the probabilities p (Fi : a) increase with n. For example, p (F1 : a) increases with n. Nevertheless, the possibilities of having the state "b" for some of these F-taxa increase with the number of taxa between A2 and A1, even if the "distance" (number of steps) between A1 and A2 do not increase.

The better possible schema is that with two known taxa in the state "a" and an out-group in the state "b".

Multistate attribute

Definition. In many cases, the attributes have more than two states (multistate attribute or character, especially for the climatic analysis). For example, in a "temperature" analysis based on the scale of DAVID *et al.* (1983), the attribute shows five states [Icy - Cold - Temperate - Hot - Torrid]. For each state, an arbitrary value can be attributed, ranging from "a" to "e". Several scenarios can be envisaged.

If the various states seem to correspond to a gradual process, it is possible to envisage the hypothesis of their polarization following a gradation without any "jump" between two successive states, from an extreme state to an other; *i.e.* [- a → b → c → d → e +] or [- e → d → c → b → a +] for a "temperature" analysis; or from an intermediate state towards the extreme states: [+ e ← d ← c (-) → b → a +] for example.

It is possible to deny any gradual process *a priori*, but to still suppose the existence of an evolution from a plesiomorphic state towards one or several apomorphic states, with possible "jumps" over some states. For example, a possible polarization would be [- a → b → c → d → e +]. In that case, the number of possible polarization's quickly increases. There are nine possibilities for three states but 64 possibilities for four states of a character.

Theoretical procedure. If it is possible to define the more probable scenario for the attribute after the analysis of the known taxa following the parsimony method : the scenario that implies the weaker quantity of homoplasies or steps. BRYANT & RUSSEL's method can be applied using similar processes of inference of the situations at the nodes as for bipolar characters. Probabilistic method can also be used with the following change : the minimal scenarios (with the lowest number of steps) that explain the known distribution of the attribute (excluding the F-taxon) are reconstructed for all the situations of the root "root a", "root b", "root c", etc.

They can be considered as equally probable, even if one can imply more steps than the other [option (1)]. The option (2) gives different weights x_a, x_b, x_c, etc., to the various situations of the roots, with the $\sum_i (x_i) = 1$. As the various values of the x_i are unknown, this option is practically inapplicable. Next step is to re-included the F-taxon in the tree, then are only accepted the situations where the alleged state for F can be reconstructed without adding supplementary homoplasies or steps in the tree ("favorable cases"), in order to keep the same minimal lengths for the new trees. This method implies that we deny any gradual process *a priori* (*contra* first scenario as above).

Application. For an attribute X with three states "a", "b" and "c" within a group of taxa, including the taxon F (Fig. 8):

BRYANT & RUSSEL's method gives the inference of "a or b or c" for F, even if the situations for A1, A2, and A3 are permuted.

Probabilistic method [with option (1)] leads to a similar conclusion but it is more precise because it will be affected by permutations of the situations for A1, A2 and A3, for examples :

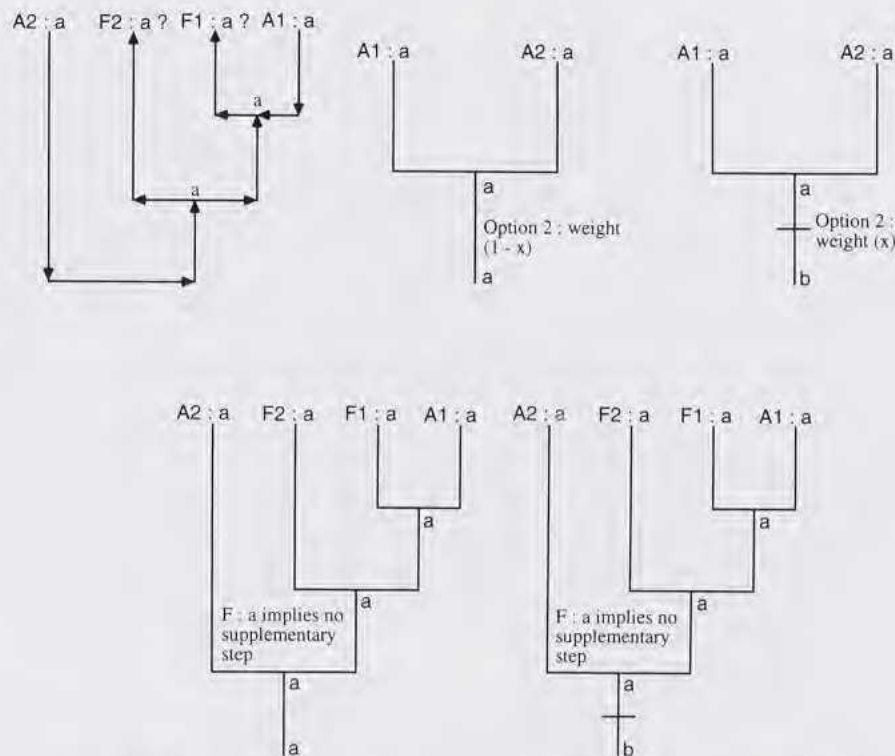If A1 is "c", A2 is "b" and A3 is "a" (Fig. 8), then p (F : c) = 2/6 = 1/3; p (F : b) = 3/6 = 1/2 and p (F : a) = 1/6.

FIG. 6. — p(F1 : a) = p(F : a) = 1 ; other situations imply supplementary steps.

If A1 is "b", A2 is "c" and A3 is "a", then p (F : b) = 2/6 = 1/3; p (F : c) = 3/6 = 1/2 and p (F : a) = 1/6.

Both these results are congruent with the position of sister-groups between A1 and F. Similar results can be obtained within each case of permutation between A1, A2 and A3.

In the situation of Figure 9, the results of the two methods slightly differ. BRYANT & RUSSEL's method gives the inference "b" for F but probabilistic method [with option (1)] gives p (F : b) = 3/4 and p (F : c) = 1/4, similarly to the situation of Figure 5.

Conclusion

It is possible to calculate a probability law for each character or attribute which is unknown for a taxon included in a phylogeny. These calculations give the maximal estimation of the probability for the inference, because the additions of steps due to the presence of the F-taxa are rejected, but they could have happened. These two different methods of inference explain the weakness of the theory of actualism concerning the ancient palaeobiotas. It is not directly the antiquity of the fossil taxa which renders less probable the inference of the attributes to the fossil, but if a fossil is older than another one, it has more "chance" to be only the sister-group of Recent taxa, thus, it only provides information of low probability. The scale of measure of the reliability of the inference is not directly temporal but phylogenetic, thus it is not directly related with the time factor.
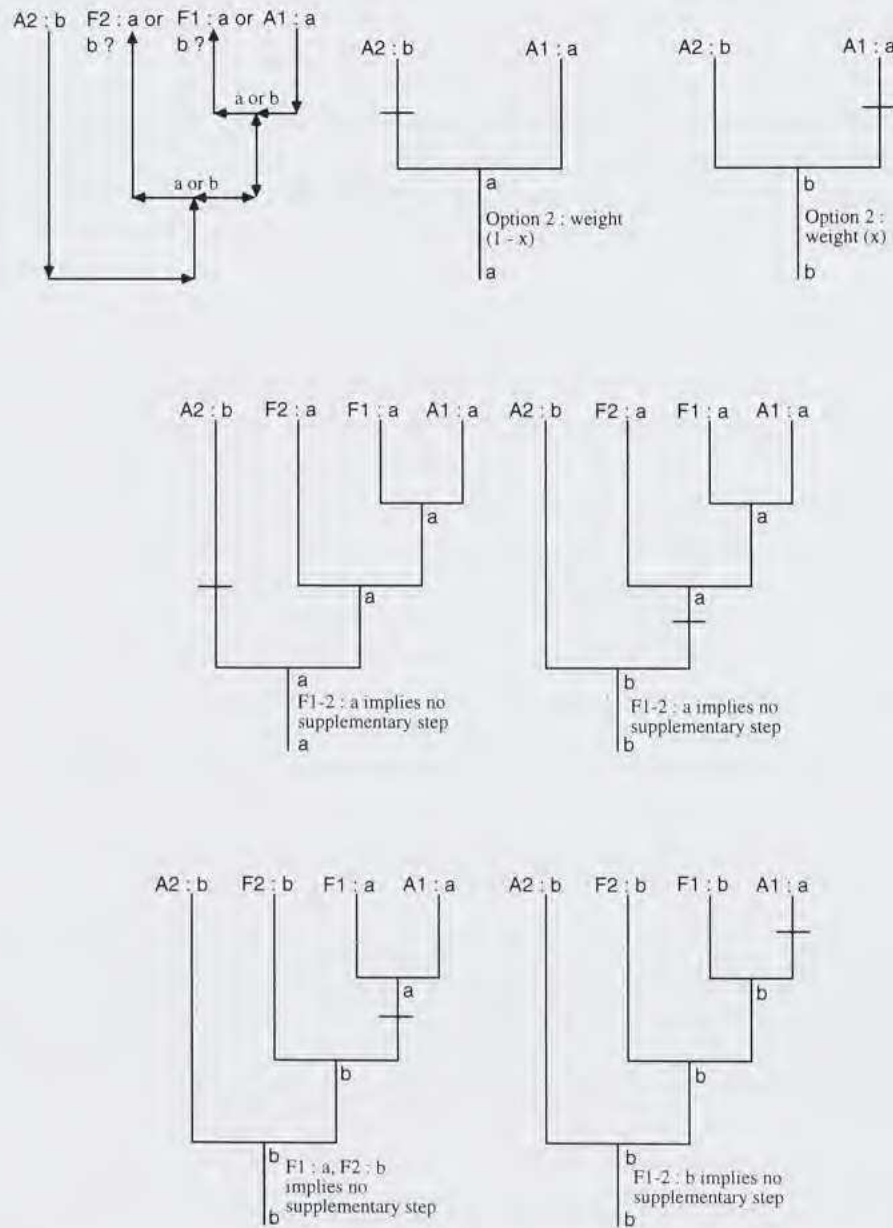
FIG. 7. — p(F1 : a) = 3/4 ; p(F2 : a) = 1/2 ; other situations also imply supplementary steps.

Application : palaeoclimatic and palaeoenvironmental phylogeny

Procedure. All the palaeoclimatological studies based on the fossil data use the comparison between fossil taxa and their "nearest" Recent relatives. More especially, palaeoclimatic studies of the Quaternary are now based on the elaborate method of the "Mutual Climatic Range" or
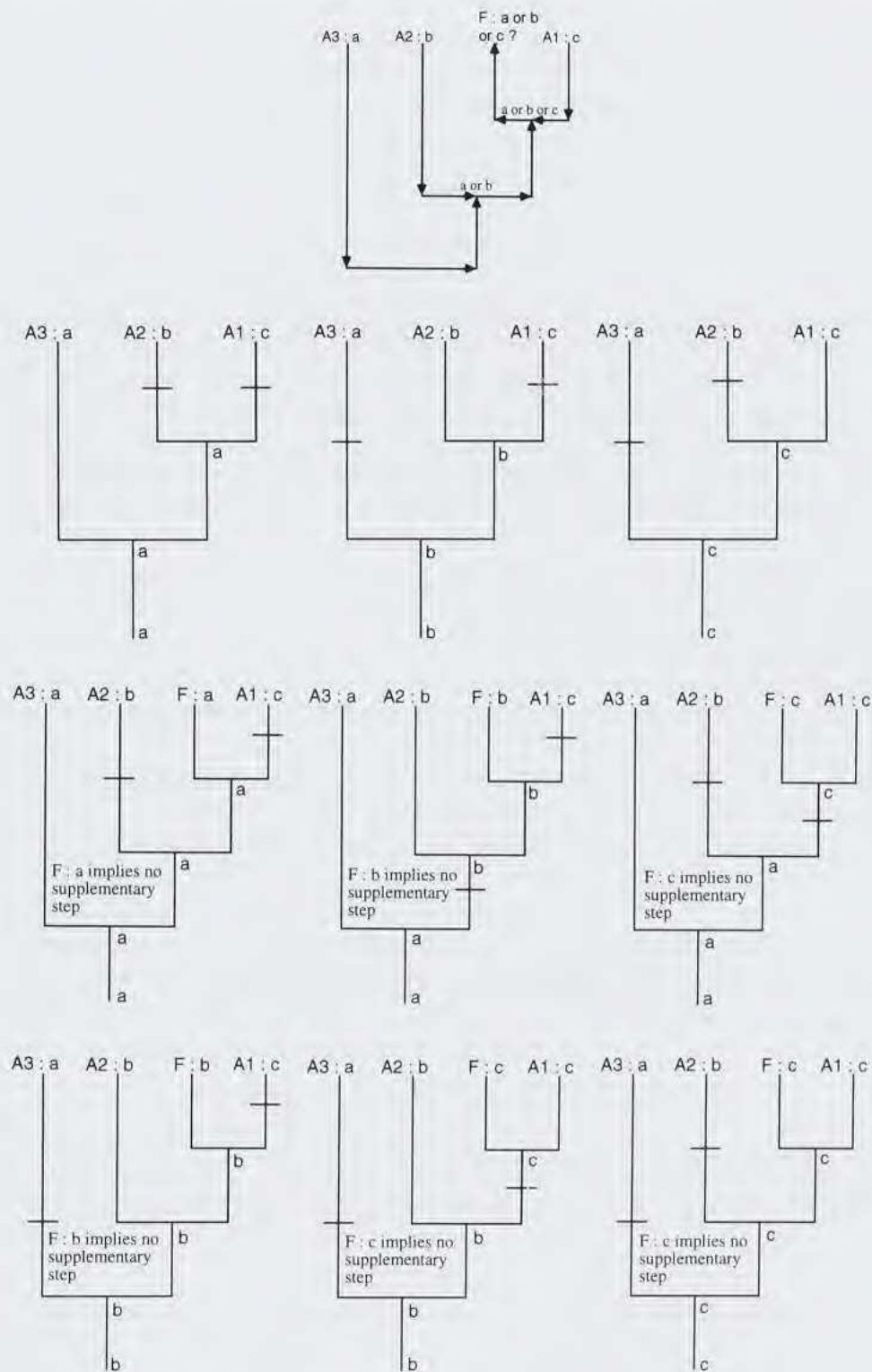
FIG. 8. — p(F : a) = 1/6 ; p(F : b) = 1/3 ; p(F : c) = 1/2.
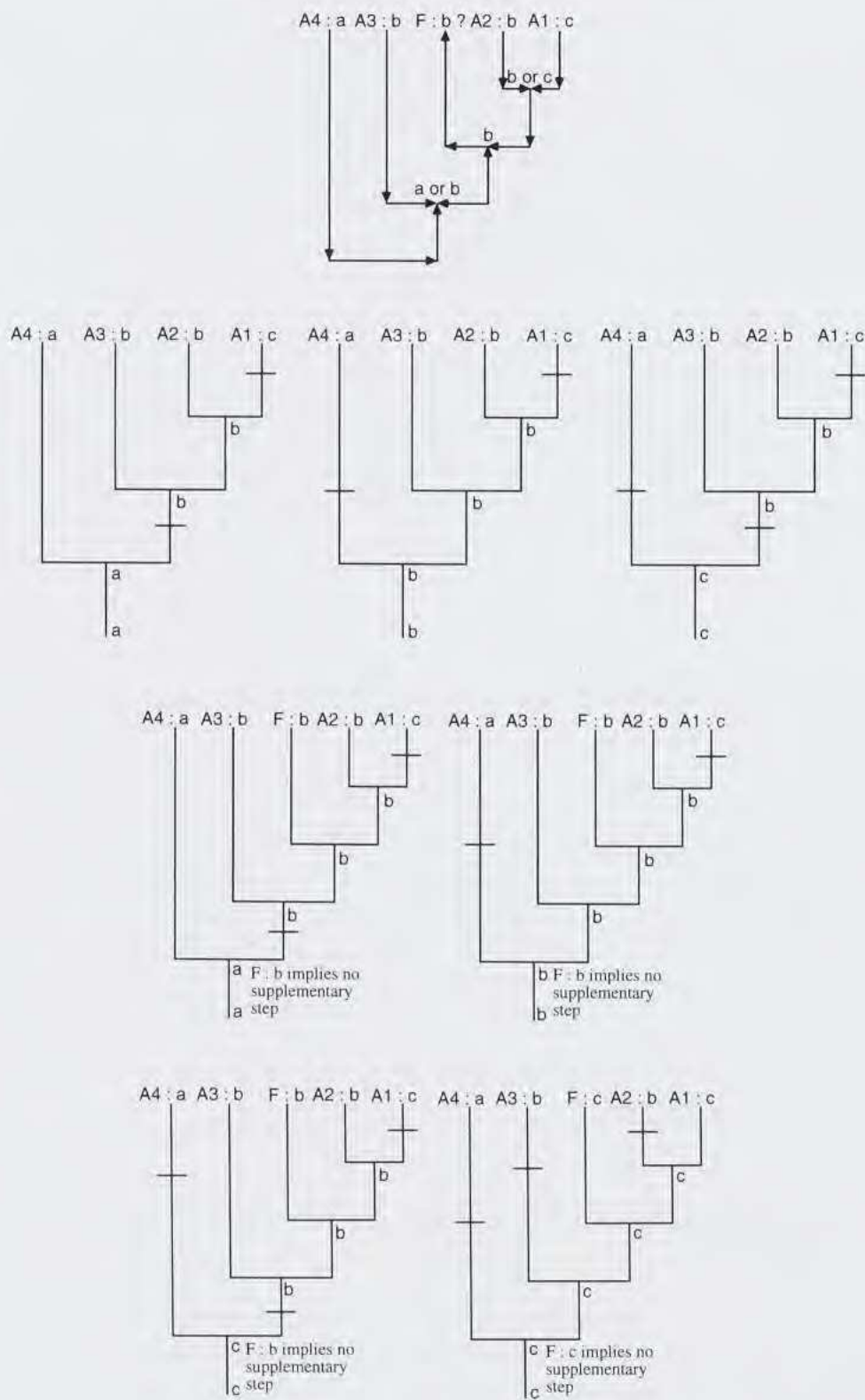
FIG. 9. — p(F : a) = 0 ; p(F : b) = 3/4 ; p(F : c) = 1/4.

"MCR" of ATKINSON *et al.* (1987). It establishes a theoretical palaeoclimate which corresponds to the "mutual intersection of the tolerance range" of the various subfossil taxa present in the studied deposit. This method implies that: the climatic tolerances of the studied species did not vary through the time, the (sub)fossil taxa can be identified as being living species, the various climatic tolerances of the (sub)fossil taxa have an intersection, all the informations share the same weight, *a priori*. This method becomes difficult to apply for strictly fossil taxa but the method of phylogenetic inference can help. The complex nature of a climate implies precise definitions of the used parameters. For example, DAVID *et al.* (1983) defined a (palaeo)climate after the combination of three types of climatic factors: [glacial - cold - temperate - hot - torrid]; [arid - dry - sub humid - humid] et [stable - alternative]. AXELROD (1992) proposed another climatic scale based on the definition of the "equable climate" characterized by a mean annual temperature of 14°C and a mean annual variation of 0°C. Whatever the scale, it is necessary to distinguish the different (palaeo)climates using discrete scales, in order to consider the data as characters (or attributes) which can be tested by a phylogenetic analysis.

The theoretical method is derived from BRYANT & RUSSELL (*loc. cit.*: 409, Fig. 1) with the two following steps:

As a first step, an analysis of inference, taxon after taxon, of the characters or attributes of unknown state. Each fossil taxon is integrated, when possible, in a phylogenetic analysis based on the present morphological characters, but not based on the attributes which shall be studied after. For each climatic attribute (temperature, humidity, stability) and each taxon, the probability law of the attribute is established. A study of correlation [structure - function] based on the preserved characters of each fossil is to be made in parallel with the study of phylogenetic inference. The conclusions of the two phylogenetic and extrapolated procedures are compared. If the results are congruent, the law of probability of the concerned attribute is taken up for the taxon. Otherwise, the taxon is considered as doubtful and is not used for the following step. Its phylogenetic placement is reexamined and its law of probability is recalculated, and compared again to the results of the study of correlation [structure - function].

As a second step, an analysis of inference of the states of the attributes for the studied palaeoenvironment is based on all the inferences established during the first step. By putting together all the data for all the taxa (F_j), for each attribute X, is calculated a series of coefficients $[P_i (X)]_i$. The "i" correspond to the states of the attribute X.

$$\text{Each } P_i (X) = \sum_j [p (X : i \text{ for } F_j)]$$

with the F_j corresponding to all the present taxa.

A law of probability L (X) of the attribute X for the concerned palaeobiota, can be established. On the basis of this law of probability, a mean value E (X) for the attribute can be calculated:

$$E (X) = P_i (X) / \sum_i [P_i (X)]$$

The results of the phylogenetic analysis are to be tested, when possible, by independent data gathered with a direct physical analysis (analysis of the Oxygen isotopes, or of Deuterium/Hydrogen, etc., MILLER *et al.*, 1988). Similarly, the results of phylogenetic biogeography are to be tested using the independent geological data (NELSON, 1985). If the results are congruent,
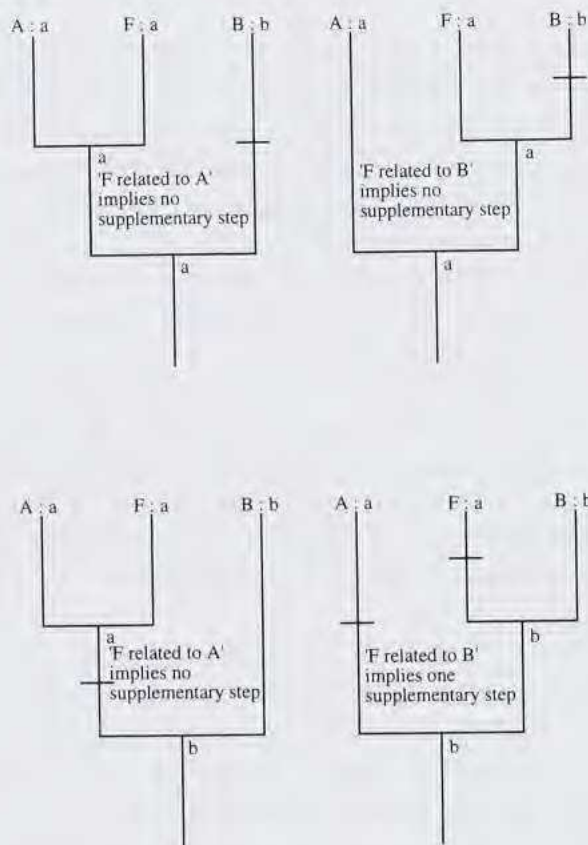
Fig. 10. — p(F is related to A) = 2/3.

they are considered as probable. Otherwise, the law of probability of the attributes can be verified for all the taxa and new data are to be found out before solving the problem.

Examples. Within this theory, for a study of palaeotemperature T (with the rate [glacial : "a" - cold : "b" - temperate : "c" - hot : "d" - torrid : "e"].

If we consider a taxon F has the following law of probability:

$$p (T : \text{"b" for F}) = 3/4, \quad p (T : \text{"c" for F}) = 1/4$$
$$\text{and } p (T : \text{"a" for F}) = 0 = p (T : \text{"d or e" for F})]$$

(corresponding to the situation in Fig. 9).

Then, F has more weight for the calculation of the law of probability of the global palaeotemperature of the palaeobiota than a taxon F' with the following law of probability for the palaeotemperature:

$$p (T : \text{"b" for F'}) = 2/6$$
$$\text{and } p (T : \text{"c" for F'}) = 3/6, \quad p (T : \text{"a" for F'}) = 1/6, \quad p (T : \text{"d or e" for F'}) = 0$$

(corresponding to the situation in Fig. 8).

This method of weighting gives a greater importance to the taxa which correspond to highly specialized climatic conditions. It allows a less empirical evaluation of the weights.

Applications concerning "Mutual Climatic Range". The same method of weighting can be applied in the subfossil record as a modification of the "Mutual Climatic Range", by giving a more important weight to the taxa specialized to only one type of biota or of climate. A Recent (or subfossil) taxon which is present under climates for which the temperature ranges from type "5" to type "3" directly gives, without inference analysis, the following law of probability:

$$p (T < 3) = 0, p (T : 3) = p (T : 4) = p (T : 5) = 1/3$$

Unlikely, a living taxon present under a climate for which the temperature is of type "5" gives the law of probability:

$$p (T < 5) = 0 \text{ and } p (T : 5) = 1$$

The same method is to be applied to all the taxa. The following analysis is that of the precedent step.

Applications concerning non inference of an attribute/allochtony. If the law of probability for an attribute of a F-taxon does not correspond to the general law of probability of the biota, its climatic or environmental constraints could have changed relatively to the nearest relative taxa or the concerned F-taxon is allochtonous for the concerned palaeobiota.

Problem of the living fossils or relic taxa (sensu DELAMARE-DEBOUTEVILLE & BOTOSANEANU, 1970). Theoretical problem : the characters or attributes which are only present in one Recent taxon A cannot be easily inferred in the fossil record, whether they are autapomorphies of A or symplesiomorphies of the group including A. For all the possible hypothesis of weighting of the homoplasies, the probability p (F : a) = p (F : b) = 1/2, situation of simple information of a sister-group relationship between F and a Recent taxon A, and the inference or the non-inference of the attribute X from A on F are equally probable. This situation occurs specially in the cases of relic taxa, which are the only Recent representatives of fossil groups. They are poorly informative in the inference of their own characters or attributes. The relic species are often located in peculiar "refuges" which are very different of the palaeobiotas of their nearest fossil relatives, as already noticed for some marine taxa which are supposed to have "migrated" from shallow water zones towards deep water zones during the Palaeozoic or Mesozoic (the Mollusk *Neopilina galatheae* Lemche, 1958 or the Crinoids for examples, DELAMARE-DEBOUTEVILLE & BOTOSANEANU, 1970).

As a first example, I can show that the Isoptera Mastotermitidae are poor palaeoclimatic indicators. The Recent termite *Mastotermes darwiniensis* (sole living representative of the family Mastotermitidae) lives under the very peculiar climate and biota of the savanna ("bush") of Northern Australia (GAY & CALABY, 1965: 396), but seems to be absent in the evergreen forest (EMERSON, 1965: 27). This insect is an excellent climatic and environmental indicator for the present days. Contrary to the hypothesis of NEL & PAICHELER (1993), the direct inference in the past of all these climatic and environmental data, using the presence of fossil *Mastotermes* spp. in Cenozoic palaeoenvironments, has a low probability of 0.5 because these fossils have (in the best case) only relationships of sister species with the Recent taxon. Thus, it is impossible to say that these fossils lived in palaeobiotas similar to that of *M. darwiniensis.* The two fossil genera *Blattotermes* and *Spargotermes*, other known Mastotermitidae, give no more palaeoclimatic and palaeoenvironmental informations because they are, in the best case, only the sister genera of *Mastotermes.* The Kalotermitidae (and other isopteran families) live under temperate, hot or

torrid climates and the Mastotermitidae live under hot or torrid climates. Either if we consider that the Mastotermitidae are really the sister-group of the Kalotermitidae (ROONWAL, 1985), or the sister-group of all other isopteran families (KAMBHAMPATI *et al.*, 1996), inference analysis of the attribute "temperature", with the two states "temperate" ("a") and "hot or torrid" ("b"), shows that the probability for the fossil Mastotermitidae to have lived under "hot or torrid" temperatures p (F : "b") = 2/3 and the contrary probability p (F : "a") = 1/3 (situation of figure 5). In a different way, BRYANT & RUSSEL'S method gives the inference of the state "b" for F and appears less precise. This example shows that only a small part of the information can be inferred in the fossil record.

As a second example, the Brachiopoda, Lingulidae show a case of ineffectiveness of the inference method. These animals usually live on the light bottom of the tide zone but PAINE (1970) indicates a species (*Lingula albida*) living in deep water. The phylogenetic inference of the former biota in the fossil record, as proposed by GALL (1971: 24) for the Triassic of Vosges (France), is difficult to establish because of the lack of phylogenetic analysis of the group which integrates the fossil taxa. The Triassic, Devonian and Ordovician Lingulidae are attributed to the genus *Lingula s.l.*, and their real relationships with the Recent genera *Glottidia* and *Lingula* remain uncertain. The probability they had lived in deep water is equal to the probability they had lived in shallow water. Some more precise informations concerning the substrate on which these animals did live can be found out after the morphological and physical analysis of the fossil shells and ancient substrate (PAINE, 1970), the phylogenetic data being useless.

INFERENCE OF THE POSITION OF A TAXON

Theoretical procedure

The available characters in the fossil record are frequently highly homoplastic (for example some of the odonatan venational structures). Thus, it is difficult to attribute a fossil taxon to a precise group on the sole basis of these ambiguous characters. Nevertheless, it is possible to estimate a probability for the event : [the taxon is related to a group rather than to another one], following a probabilistic method similar to the precedent. This method can be applied to a taxon F clearly related to two possible groups A and B, but which does not share any clear synapomorphy with A or B, to the exclusion of one of the two groups. The two events [F is related to A] and [F is related to B] are opposite.

Two hypotheses of scenario can be made:

First, with the supplementary hypothesis that, for each concerned character, the probabilities for the additions of new steps are equal to zero and that the other possible situations of polarity are equally probable, the probability of the events [F is related to A] and [F is related to B] can be calculated with the quotient of the number of favorable cases by the total number of possible cases.

Second, a weight p (arbitrary, medium or maximal) can be given to the simple addition of a new step (0 < p < 1 if p is calculated as a percentage or a rate of homoplasy). The non-addition of a step will have a weight q = 1 − p. Furthermore, [p = 0] corresponds to the probability zero for the addition of a new step.

X is a character that is supposed to have two states "a" and "b". As the polarity of the character is completely unknown, we consider that it can equally be in the states "a or b" in the root of the tree. Then, in the simple tree made with the two taxa A and B, the minimal scenarios

(with the lowest number of steps) that explain the known distribution of the attribute (excluding the F-taxon) are reconstructed for the two situations "root a" and "root b", and considered as equally probable, even if one can imply more steps than the other. After, the F-taxon is re-included in the partial tree of the taxa A and B, then are only accepted the situations where the alleged state for F can be reconstructed without adding supplementary homoplasies or steps in the tree ("favorable cases"), in order to keep the same minimal lengths for the new trees.

Examples

If the state "a" is shared by F, A and B, no homoplastic situation appears whether we consider X as an apomorphy or plesiomorphy. The character "a" is not informative.

$$p\ ([F\text{ is related to A}]) = p\ ([F\text{ is related to B}]) = 1/2$$

If the state "a" is present in F and A but absent in B, the polarity of which is ambiguous (presence of several homoplastic situations concerning the character in the phylogenetic analysis) (Fig. 10). If $p = 0$, there are only three "possible" situations, with two "favorable" for the hypotheses [F is related to A], the universe of the possibilities is: {(root a ; F is related to A) noticed (root a ; F-A) ; (root a ; F is related to B) noticed (root a ; F-B) ; (root b ; F is related to A) noticed (root b ; F-A)}. F is more probably to related to A than B.

Probabilities are:

$$p\ (F\text{ is related to A}) = 2/3 \text{ and } p\ ([F\text{ is related to B}]) = 1/3$$

If we suppose that p for the added steps is not nil, there are four "possible" situations with different weight. The universe of the possibilities is: {(root a ; F-A) with a weight q ; (root a ; F-B), q ; (root b ; F-A), q ; (root b ; F-B) p}.

The probability are:

$$p\ ([F\text{ is related to A}]) = (2q)\ /\ (3q + p)\text{ or}$$
$$p\ ([F\text{ is related to A}]) = (2q)\ /\ (2q + 1)$$

As there are two favorable cases with the weight q against four possible cases, including three cases with a weight q and one with the weight p. If $p = 0$, we find again $q = 1$ and the precedent probability. For all the possible values of p, the "best" possible value of the probability is that corresponding to $p = 0$, because $2/3 > (2q)\ /\ (3q + p)$ for all values of p :

$$(2q)\ /\ (3q + p) = 2\ (1 - p)\ /\ (3(1 - p) + p) = 2\ (1 - p)\ /\ (3 - 2p) = (2 - 2p)\ /\ (3 - 2p)$$
$$\text{and } (2 - 2p)\ /\ (3 - 2p) < 2/3 \text{ if } (2 - 2p) \times 3 < (3 - 2p) \times 2 \text{ if } 6 - 6p < 6 - 4p$$
$$\text{if } 6p > 4p \text{ (with } 0 < p < 1)$$

If we add a character X2, also of uncertain polarity, independent of X1 and with state "a" shared by F and A and state "b" for B, the universe of the possible events is the product of the universes of the possibilities corresponding to X1 and to X2.

If the weight $p = 0$ for the homoplasies, the universe is: {(root "a" for X1, root "a" for X2 and F is related to A) [or with an abbreviated notation: (a: X1, a: X2 ; F-A)] ; (a: X1, b: X2 ; F-A) ; (b: X1, a: X2 ; F-A) ; (b: X1, b: X2 ; F-A) ; (a: X1, a: X2 ; F-B)}. All the cases of the type (one of the roots is in the state "b" and [F is related to B]) imply an homoplasy and are not counted. There are five possible cases with four favorable to [F is related to A]. Consequently, p ([F is related to A]) = 4/5 and p ([F is related to B]) = 1/5. The probability for (F is related to A) increases.

If the weight p is not nil, the homoplasies are counted and the universe becomes $\{$(a: X1, a: X2 ; F-A), weight q^2 ; (a: X1, b: X2 ; F-A), q^2 ; (b: X1, a: X2 ; F-A), q^2; (b: X1, b: X2 ; F-A), q^2; (a: X1, a: X2 ; F-B), q^2; (b: X1, a: X2 ; F-B), pq ; (a: X1, b: X2 ; F-B), pq ; (b: X1, b: X2 ; F-B), $p^2\}$. There are four cases favorable to "F-A" with a weight q^2. There are also four cases unfavorable to "F-A", one with a weight p^2, two with a weight pq and one with a weight q^2.

$$p ([F \text{ is related to } A]) = (4\ q^2) / (5\ q^2 + 2\ pq + p^2) = (4\ q^2) / [4q^2 + (p + q)^2]$$

As (p + q) = 1, I find:

$$p ([F \text{ is related to } A]) = (4\ q^2) / [4q^2 + 1]$$

If p = 0, q = 1 we find again p ([F is related to A]) = 4/5. The best possible value of p ([F is related to A]) occurs when p = 0 (the homoplasies are impossible) because : $(4\ q^2) / [4q^2 + 1] < 4/5$ is equivalent to $4q^2 < 4$ or $q^2 < 1$. Otherwise, if p is different of 0, the probability p ([F is related to A]) varies between 4/5 and 0.

In the case of n characters X1, X2, X3, X4, ..., Xi,..., Xn which are all in the same situation (with an uncertain polarity, independent, with state "a" shared by F and A and state "b" for B),

If p = 0, the universe of the possibilities is of cardinal $(2^n + 1)$, with (2^n) events in favor of [F is related to A], thus:

$$p ([F \text{ is related to } A]) = (2^n) / (2^n + 1) \text{ and } p ([F \text{ is related to } B]) = 1 / (2^n + 1)$$

If p is not nil, the cardinal of the universe increases to the value $2 \times 2^n = 2^{n+1}$.

2^n events are favorable to [F is related to A] with a weight q^n, the other events correspond to [F is related to B], with one having the same weight q^n, n events have the weight pq^{n-1}, $(n!) / [2! (n-2)!]$ events have the weight $p^2 q^{n-2}$, $(n!) / [3! (n-3)!]$ events have the weight $p^3 q^{n-3}$, etc., and one event has the weight p^n.

The probability $p ([F \text{ is related to } A]) = (2^n q^n) / [(2^n + 1)q^n + npq^{n-1} + \{(n!)/[2!(n-2)!]\}p^2q^{n-2} + \{(n!)/[3!(n-3)!]\}p^3q^{n-3} + ... + p^n]$. There is an usual remarkable identity in the denominator thus:

$$p ([F \text{ is related to } A]) = (2^n q^n) / [(2^n q^n + 1)]$$

This formula generalizes the preceding ones. Furthermore, if p = 0 and q = 1, we find again the formula $(2^n) / (2^n + 1)$. For all cases, the maximal value of p ([F is related to A]) is equal to $(2^n) / (2^n + 1)$, when p varies from 0 to 1.

If there is a character X (bipolar, a or b) with the state "a" shared by A and F but not by B and one character Y (bipolar, a or b) with the state "b" shared by B and F but not by A (situation symmetrical of Figure 10).

In the case of a weight p = 0 for the added steps, the universe of the possibilities is : $\{$(a : X, a : Y ; F-A) ; (b : X, a : Y ; F-A) ; (a : X, b : Y ; F-B) ; (a : X, a : Y ; F-B)$\}$; there are four events with two favorable to [F is related to A] and 2 are favorable to [F is related to B], p ([F is related to B]) = p ([F is related to A] = 2/4 = 1/2.

If p is not nil for the added steps, the universe becomes : $\{$(a : X, a : Y ; F-A), q^2 ; (b : X, a : Y ; F-A), q^2 ; (a : X, b : Y ; F-B), q^2 ; (a : X, a : Y ; F-B), q^2 ; (a : X, b : Y ; F-A), pq ; (b : X, a : Y ; F-A), pq ; (b : X, a : Y ; F-B), pq ; (a : X, b : Y ; F-B), pq$\}$. p ([F is related to A]) = $(2q^2 + 2\ pq) / (2q^2 + 2\ pq + 2q^2 + 2\ pq) = 1/2$.

In the two hypothesis, the two informations of X and Y "neutralize" each other.

Generalization

If we have n characters of the type X with the state "a" in common to F and A and m characters of the type Y with the state "b" in common to F and B.

In the case of a weight p = 0 for the added steps, the universe holds $[(2^n) + (2^m)]$ events which are distributed into (n+ m) -upsets of two types:

(2^n) events of the type (a : Y1, a : Y2, ..., a : Yi, ..., a : Ym, a or b : X1, a or b : X2, ..., a or b : Xi,..., a or b : Xn, F-A);

(2^m) events of the type (a or b : Y1,..., a or b : Yi, ..., a or b : Ym, a : X1, ..., a : Xi, ..., a : Xn, F-B).

Consequently,

$$p ([F \text{ is related to } A]) = (2^n) / [2^n + 2^m]$$
$$p ([F \text{ is related to } B]) = (2^m) / [(2^n) + (2^m)]$$

In the case of a weight p for added steps, then:

$$p ([F \text{ is related to } A]) = 2^n q^n [q^m + mpq^{m-1} + ... + p^m] / 2^n q^n [q^m + mpq^{m-1} + ... + p^m] + 2^m q^m [q^n + npq^{n-1} + ... + p^n],$$

then:

$$p ([F \text{ is related to } A]) = 2^n q^n [p + q]^m / (2^n q^n [p + q]^m + 2^m q^m [p + q]^n)$$

as (p + q) = 1, then:

$$p ([F \text{ is related to } A]) = 2^n q^n / (2^n q^n + 2^m q^m)$$

This formula generalizes and replaces all the precedent ones.

p = 0 and q = 1 give again p ([F is related to A]) = $(2^n) / [2^n + 2^m]$ which is the maximal possible value when p varies from 0 to 1. If m = n, p ([F is related to A]) = 1/2. The contradictory informations "neutralize" each other.

Furthermore, p ([F is related to A]) = 1/2 (for all the values of m and n) if q = 1/2 *i.e.* if the weight p of the addition of steps = 1/2. Even if there are distinctly more characters in favor of a relation with A rather than with B (n >> m), if the probability that all these characters implies additions of new steps is too important, it is impossible to decide.

BECHLY *et al.* (1997) apply this method to the peculiar case of the Lower Cretaceous English Zygoptera *Cretacoenagrion* (taxon of uncertain position because of the lack of information). There is a maximal probability of 4/5 for the event [*Cretacoenagrion* is related to the Lestoidea rather than to the Coenagrionoidea] but it is still impossible to state positively that it is a Lestoidea.

If the number of shared characters between F and A but not by B increases, the probability of the event [F and A are related] increases. This result is congruent with an intuitive approach of the problem. This method does not prove that F is really related with A and would not replace the cladistic method based on the principle of the fundamental importance of the synapomorphies for the determination of the relationships between the taxa. This method gives an estimate of the probability p ([F is related to A]) but the calculation of the exact value of this probability depends on the determination of the rate p of the homoplasies. The result can greatly vary with the value of p. A probability, even very high, is not a proof.

CONCLUSION

Although these methods of probabilistic inferences could appear not very easy to use, they have the advantage of quantifying the possibilities of transferring Recent biological and environmental data to (sub)fossil taxa. Thus, they limit and refine possibilities of global transferring of actualism. Quantification of inferred palaeoclimatic data allows establishments of more precise palaeoclimatic hypotheses, susceptible of being tested by physical analysis. Comparisons between palaeoclimatic hypotheses of different palaeobiotas shall be easier to attempt because these hypotheses are based on the same method. The probabilistic inferences of taxa positions cannot replace phylogenetic analyses but they are better than subjective and not quantified hypotheses.

ACKNOWLEDGEMENTS

REFERENCES

ATKINSON, T. C., BRIFFA, K. R. & COOPE, G. R., 1987. — Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature*, **325**: 587-592.

AXELROD, D. I., 1992. — What is an equable climate? *Palaeogeography, Palaeoecology, Palaeoclimatology*, **91**: 1-12.

BECHLY, G., MARTINEZ-DELCLOS, X., JARZEMBOWSKI, E. A., NEL, A. & ESCUILLIE, F, 1997. — The Mesozoic non-calopterygoid Zygoptera : description of new genera and species from the Lower Cretaceous of England and Brazil and their phylogenetic significance (Odonata, Zygoptera, Coenagrionoidea, Hemiphlebioidea, Lestoidea). *Cretaceous Research*, **17**: in press.

BRYANT, H. N. & RUSSELL, A. P., 1992. — The role of phylogenetic analysis in the inference of unpreserved attributes of extinct taxa. *Philosophical Transactions of the Royal Society of London, B*, **337**: 405-418.

DARLU, P. & TASSY, P., 1993. — *Reconstruction phylogénétique, concepts et méthodes*. Paris, Masson: 1-245.

DAVID, B., DAVID, L., DURAND, J., EISENMANN, V., FISCHER, J.-C., FRANIATTES, S., GALL, J.-C., GOUJET, D., HEYLER, H., LAURIN, B., LE RENARD, J., MERLE, D., PHILIP, J., PLAZIAT, J.-C., RIOULT, B., ROUX, M., TAQUET, P. & THIERRY, J., 1983. — Proposition d'une typologie critique des facteurs de paléomilieux (paléomilieux continentaux et marins non profonds). *Document. du Premier Congrès international de Paléoécologie*, Lyon, 18-23 JUILLET 1983: 1-13.

DELAMARE-DEBOUTEVILLE, C. & BOTOSANEANU, L., 1970. — *Formes fossiles vivantes*. Paris, Hermann: 1-232.

DELEPORTE, P., 1993. — Characters, attributes and tests of evolutionary scenarios. *Cladistics*, **9**: 427-432.

DE PINNA, M. C. C., 1991. — Concepts and tests of homology in the cladistic paradigm. *Cladistics*, **7**: 367-394.

EMERSON, A. E., 1965. — A review of the Mastotermitidae (Isoptera), including a new fossil genus from Brazil. *American Museum Novitates*, **223**: 1-45.

FURON, R., 1964. — Paléoclimatologie, paléomagnétisme et dérive des continents. *Compte Rendu Sommaire des Séances de la Société de Biogéographie*, Paris, 40ème année, **351-355** : 87-91.

GALL, J.-C., 1971. — Faunes et paysages du grès à Voltzia du nord des Vosges. Essai paléoécologique sur le Buntsandstein supérieur. *Mémoires du Service de la Carte géologique d'Alsace et Lorraine*, Strasbourg, **34**: 1-318.

GAY, F. J. & KALABY, J. H., 1970. — Termites of the Australian region, pp. 393-448. *In* : K., KRISHNA & F. M., WEESNER, *Biology of Termites, volume 2*. New York, Academic Press: 1-643.

GRANDCOLAS, P., 1993. — The origin of biological diversity in a tropical cockroach lineage : a phylogenetic analysis of habitat choice and biome occupancy. *Acta Œcologica*, **14**: 259-270.

GRANDCOLAS, P., DELEPORTE, P. & DESUTTER-GRANDCOLAS, L., 1994. — Why to use phylogeny in evolutionary ecology? *Acta Oecologica*, **15**: 661-673.

KAMBHAMPATI, S., KJER, K. M. & THORNE, B. L., 1996. — Phylogenetic relationship among termite families based on DNA sequence of mitochondrial 16S ribosomal RNA gene. *Insect Molecular Biology*, **5**: 229-238.

MADDISON, W. P., DONOGHUE, M. J. & MADDISON, D. R., 1984. — Outgroup analysis and parsimony. *Systematic Zoology*, **33**: 83-103.

MICKEVICH, M. F. & WELLER, S. J., 1990. — Evolutionary character analysis : tracing character change on a cladogram. *Cladistics*, **6**: 137-170.

MILLER, R. F., FRITZ, P. & MORGAN, A. V., 1988. — Climatic implications of D/H ratio in beetle chitin. *Palaeogeography, Palaeoecology, Palaeoclimatology*, **66**: 277-288.

NELSON, G., 1985. — Plate tectonics and biogeography. *Journal of History of Earth Society*, **4**: 187-196.

NEL, A. & PAICHELER, J.-C., 1993. — Les Isoptera fossiles. Etat actuel des connaissances, implications paléoécologiques et paléoclimatiques. *Cahiers de Paléontologie, CNRS*: 102-179.

PAINE, R. T., 1970. — The sediment occupied by recent lingulid Brachiopds and some palaeoecological implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **7**: 21-31.

PLATNICK, N. I., GRISWOLD, C. E. & CODDINGTON, J. A., 1991. — On missing entries in cladistic analysis. *Cladistics*, **7**: 337-343.

ROONWAL, M. L., 1985. — Recent researches on wing microsculpturing in Termites (Isoptera), and its evolutionary and biological significance. *Proceedings of the Indian National Science Academy, B*, **51**: 135-168.

SWOFFORD, D., 1991. — *Paup Version 3.0s*. Champaign, Illinois Natural History Survey: 1-178.